Molecular Genetic Analysis of U.S. and Chinese Soybean Ancestral Lines

Zenglu Li, Lijuan Qiu, Jeffery A. Thompson, Molly M. Welsh, and Randall L. Nelson*

ABSTRACT

Most of the U.S. soybean [*Glycine max* (L.) Merr.] ancestral lines were introduced from China, but nothing is known of the genetic relationships among the ancestors of modern U.S. and Chinese cultivars. The objectives of this research were to measure the variation among the major ancestors of U.S. and Chinese cultivars, to establish the genetic relationships among these U.S. and Chinese soybean ancestral lines, and to determine the relationship between geographical origin and genetic diversity. Genomic DNA from these lines was characterized by random amplified polymorphic DNA (RAPD) with 35 selected decamer primers. On the basis of the presence or absence of amplified DNA fragments, simple matching coefficients were used to calculate genetic similarities between pairs of lines. Cluster analyses generally separated the ancestral gene pools of the USA and China. Clusters reflected the geographical origin of the lines. Large differences exist between northern U.S. and Chinese ancestral lines and central and southern Chinese ancestral lines. The pattern of diversity found within the U.S. and Chinese ancestors can aid breeders in selecting parental lines to more efficiently exploit the diversity found in these two major gene pools.

Soybean originated in China and is a crop of major importance in both China and the USA. Both countries have active soybean breeding programs. Over 650 cultivars were released in China from 1923 to 1995 (Cui et al., 1999). These cultivars were derived from 348 soybean ancestral lines including 302 Chinese landraces, 24 U.S. cultivars, 12 Japanese cultivars, and 10 cultivars from other countries (Cui et al., 1999). There have been over 400 publicly released cultivars in the USA, which were developed from approximately 80 soybean ancestral lines (Gizlice et al., 1994). Although most of the ancestors of U.S. soybean cultivars originally came from China in the early part of the 20th century, the genetic relationship between the two ancestral gene pools that produced the current cultivars of the USA and China is unknown. The genetic base of soybean breeding in North America is very limited (Gizlice et al., 1994; Sneller, 1994). Twenty-eight introductions and seven first progenies (U.S.-developed cultivars with uncertain pedigrees) have contributed over 95% of the genes in publicly released cultivars (Gizlice et al., 1994). The cultivars from China could be a potentially important resource for genetic diversity for many traits such as disease resistance, seed composition, and yield to improve U.S. cultivars. As more cultivars are exchanged between China and the USA, an understanding of the genetic diversity in the ancestral lines of these cultivars will aid plant breeders in selecting parents to enhance the performance of future soybean cultivars.

The objectives of this research were (i) to measure the variation among the major ancestors of the U.S. and Chinese cultivars; (ii) to establish the genetic relationships among these U.S. and Chinese soybean ancestral lines; and (iii) to determine the relationship between geographical origin and genetic diversity.

MATERIALS AND METHODS

Plant Materials

On the basis of ecological conditions and cropping systems, there are three major soybean production regions in China. They are the northeast (NE), the Huang Huai Hai region (HHH) in east central China, and the south (Zhang, 1985), which account for approximately 40, 35, and 20% of soybean planting area in China, respectively. On the basis of the contribution of ancestral lines to released cultivars, 32 important soybean ancestors (or selections from soybean ancestors) were chosen to represent these three major soybean production regions (Table 1). Seeds of the Chinese ancestral lines were not explicitly preserved. To identify extant lines that are most likely to be the same genotypes as those used in the original crosses, we consulted many sources in the Chinese literature to learn not only the history of Chinese cultivars but also the descriptions of parental lines. Since the names given to many of these ancestral lines are not unique, it is possible that in some cases the wrong genotype may have been included in this research even though the name matches that presented in the pedigree. The ancestral lines chosen for this research are in the pedigrees of more than 75% of all Chinese cultivars released during the past 75 yr (Anonymous, 1980; Zhang, 1985; Chang and Sun, 1991; Hu and Tian, 1993; Cui et al., 1999). More cultivars have been released in the northeast so more ancestors were selected from the northeast than from the HHH or the south (Table 1). The most important ancestors from the northeast occur in the pedigrees of more than 200 cultivars. In the HHH, none of the ancestors contributed to more than 61 cultivars, and in the south the greatest contributing ancestor occurs in the pedigree of only 20 cultivars (Table 1). The ancestors that originated in each region have made the greatest contribution to cultivars developed in that region. The ancestors from NE China occur in the pedigrees of 82% of the cultivars released in the northeast. The ancestors from

Z. Li, Dep. of Crop Sciences, 1101 W. Peabody Dr., Univ. of Illinois, Urbana, IL 61801; L. Qiu, Institute of Crop Germplasm Resources, Chinese Academy of Agricultural Sciences, Beijing, China; J.A. Thompson, Pioneer Hi-Bred Intl., P.O. Box 328, Hamel, IL 62046; M.M. Welsh, USDA-ARS, Plant Germplasm Introduction and Testing Research, Room 59 Johnson Hall, Washington State Univ., Pullman, WA 99164-6402; R.L. Nelson, USDA-Agricultural Research Service, Soybean/Maize Germplasm, Pathology, and Genetics Research Unit, Dep. of Crop Sciences, 1101 W. Peabody Dr., Univ. of Illinois, Urbana, IL 61801. Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the USDA or the University of Illinois and does not imply its approval to the exclusion of other products or vendors that may also be suitable. Received 5 June 2000. *Corresponding author (rlnelson@uiuc.edu).

Table 1. Major Chinese soybean ancestral lines selected for diversity analysis.

PI number†	Ancestral line	Region	Province	Number of cultivars derived	Code#	Note
PI578497A	Jin yuan	NE‡	Liaoning	243	A142	
PI578493	Huang bao zhu	NE	Jilin	217	C333	A selection from Si li huang (A203)
PI561354	Zi hua No. 4	NE	Heilongjiang	130	C287	A selection from Bai mei (A019)
PI458506	Feng di huang	NE	Jilin	92	C324	A selection from Du lu dou (A071)
PI578503	Tie jia si li huang	NE	Jilin	89	A219	
PI602502	Xiong yue xiao huang dou	NE	Liaoning	58	A240	
PI602497	Ke shan si li jia	NE	Heilongjiang	57	A149	
PI464916	Ji ti No. 2	NE	Liaoning	29	C485	A selection from Tie jia zi (A220)
PI458510	Ji ti No. 1	NE	Liaoning	28	C484	A selection from Xiao jin huang (A235)
PI602498	Xiao jin huang	NE	Jilin	27	A234	
PI458505	Da bai mei	NE	Liaoning	22	A040	
PI464917	Ji ti No. 3	NE	Jilin	20	C368	A selection from Si li huang (A204)
PI297505	Ji ti No. 5	NE	Heilongjiang	20	C370	A progeny from Hai lun jin yuan (A104), Da bei mei (A040), and Huang bao zhu (C333)
PI430595	58-161	HHH§	Jiangsu	61	C417	A selection from Bin hai da bai hua (A033)
PI602501	Tong shan tian e dan	HHH	Jiangsu	61	A224	
PI468408A	Qi huang No. 1	HHH	Shandong	58	C555	A selection from A288 (Unknown)
PI578498B	Ju xuan 23	HHH	Shandong	54	C540	A selection Ji mo you dou (A129)
PI567604A	Xin huang dou	HHH	Shandong	52	C574	A selection from Yi du ping ding huang (A247)
PI602499	Tie jiao huang	HHH	Shandong	49	A221	
PI602991	Shandong si jiao qi	HHH	Shandong	15	A189	
PI602993	Pi xian ruan tiao zhi	HHH	Jiangsu	14	A170	
PI602992	Qin yang shui bai dou	HHH	Henan	14	A183	
PI578491A	Hua xian da lu dou	HHH	Henan	12	A118	
PI578495	Jin dou No. 4	HHH	Shanxi	9	C594	A progreny from Ji zao huang (A132) and Shandong xiao huang dou (A190)
PI578488B	Feng xian sui dao huang	South¶	Shanghai	20	A081	
PI464932	Nan nong 493-1	South	Jiangsu	18	C431	A selection from 51-83 (A002)
PI436562	Ai jiao zao	South	Hubei	10	C288	A selection from A284 (Unknown)
PI578499A	Shanghai liu yue bai	South	Shanghai	9	A194	
PI602994	Pu dong da huang dou	South	Shanghai	5	A176	
PI32454	Tai xing hei dou	South	Jiangsu	5	A212	
PI430620	Hou zi mao	South	Hubei	5	A115	
PI578504	Xiang dou No. 3	South	Hunan	4	C307	A selection from Shao dong liu yue huang (A198)

† Plant introduction.
‡ Northeast region in China.
§ Huang Huai Hai region in east central China.
¶ Southern region in China.
# The codes correspond with those in Cui et al. (1999).

the HHH and southern regions are in the pedigrees of approximately 57 and 42%, respectively, of the cultivars released in these regions (Table 2).

The Chinese ancestral lines included in this study are not the same as those defined as the major ancestors by Cui et al. (2000a). The work of Cui et al. (2000a) was based solely on pedigrees without consideration if the parental lines were still extant and available. All entries in this research were initially available in the USDA Soybean Germplasm Collection or were obtained from the Institute of Crop Germplasm Resources, Chinese Academy of Agricultural Sciences, Beijing, China. Many early cultivars developed in China were direct selections from original ancestors. Thirteen Chinese lines included in this study were such selections (Table 1). In many cases, the sole or major contribution of ancestral lines was made through these selections so characterization of these selections provides more accurate information about the genetic contribution of these ancestral lines. Two Chinese cultivars included in this research were developed by hybridization and included the contribution of five ancestral lines (Table 1). The Chinese ancestral lines or direct descendants of Chinese ancestral lines included in the study represent 19 of top 22 genetic contributors to released Chinese cultivars and account for over 40% of the genes in Chinese soybean breeding gene

pool as defined by Cui et al. (2000a). Table 1 includes identification codes for Chinese lines that can be used to reference these lines in the USDA Technical Bulletin 1871 (Cui et al., 1999).

A combination of 18 soybean ancestors and first progenies, which were defined by Gizlice et al. (1994) and represent more than 85% of the genetic base of North American soybean cultivars, were selected to represent the U.S. ancestors in this research (Table 3). All accessions used in this research are in the USDA Soybean Germplasm Collection (Urbana, IL).

Table 2. The percentage of soybean cultivars released from each region that contain selected ancestors from each region in their pedigrees.

	Region of origin of released cultivars		
Origin of selected ancestors	Northeast†	HHH‡	South§
	%		
Northeast	82	25	4
HHH	1	57	22
South	0	5	42

† Northeast region in China.
‡ Huang Huai Hai region in east central China.
§ Southern region in China.

Table 3. Major contributing U.S. soybean ancestors or first progeny selected for diversity analysis and the percentage of genes theoretically contributed to U.S. cultivars†.

Ancestor	All cultivars	Northern cultivars	Southern cultivars	Origin
	%			
Lincoln	17.9	24.2	2.9	‡
Mandarin (Ottawa)	12.2	17.2	0.0	China
CNS	9.4	3.0	24.7	China
Richland	8.2	11.3	0.8	China
S-100	7.5	1.8	21.3	China
Ogden	4.9	4.3	6.4	‡
A.K. (Harrow)	4.9	6.9	0.0	China
Dunfield	3.6	3.5	3.9	China
Mukden	3.5	4.9	0.0	China
Jackson	3.3	0.2	10.6	‡
Illini	2.2	3.1	0.04	China
Perry	2.1	2.1	2.1	‡
Roanoke	2.1	0.2	6.5	China
Capital	1.7	2.4	0.0	China
Haberlandt	0.8	0.1	2.5	N. Korea
Ralsoy	0.6	0.1	1.9	N. Korea
Arksoy	0.5	0.04	1.7	N. Korea
Korean	0.5	0.8	0.0	N. Korea
Total	85.9	86.1	85.3	

† Data from Gizlice et al. (1994), used with permission of the authors.
‡ First progenies are U.S. developed cultivars with uncertain parent(s).

DNA Isolation and RAPD Assay

Genomic DNA was isolated from the fresh unifoliolate leaf tissue of ten greenhouse-grown plants for each genotype. The CTAB (cetyltrimethylammonium bromide) method of Keim et al. (1988) with some modifications was used for DNA isolation. Leaves from each sample were ground in liquid nitrogen, and 700 μL of CTAB buffer [1.4 *M* NaCl; 100 m*M* Tris pH 8.0; 2% (w/v) CTAB; 20 m*M* EDTA; 0.5% (w/v) Na bisulfate and 1% (v/v) 2-mercaptoethanol] were added to suspend the powdered materials. The samples were incubated in a water bath at 65°C for 1 h and then 500 μL chloroform/isoamyl alcohol (24:1, v/v) were added. After shaking for 1 h at room temperature, the samples were spun at 15 000 rpm (Beckman Microfuge E, Beckman Coulter, Fullerton, CA) for 5 min at 4°C. The supernatant was transferred to a new 1.5-mL tube with 2 μL RNase (2 mg/mL) and then incubated at 37°C for 1 h. Four-fifth volume of isopropanol was added to each tube and the tubes were centrifuged at 15 000 rpm for 5 min. The supernatant was decanted and the DNA pellet was washed with 70% (v/v) ethanol. The DNA pellets were then dried and dissolved in 100 μL TE buffer. Total genomic DNA was standardized to a uniform concentration (10 ng/μL) with a Perkin Elmer UV/VIS spectrometer (Perkin-Elmer Corporation, Norwalk, CT) for the polymerase chain reaction (PCR) reactions.

Thirty-five decanucleotide primers obtained from Operon Technologies (Alameda, CA) and selected for high diversity scores among diverse soybean lines (Thompson and Nelson, 1998) were used in this research. The RAPD protocol reported by Kresovich et al. (1994) was followed with minor modifications, and the DNA was amplified in a Perkin-Elmer Gene-Amp PCR System 9600. A 25 μL reaction volume was used with 50 ng of genomic DNA. The amplification program consisted of 2 min at 94°C followed by 45 cycles of 1 min at 94°C, 5 min at 38°C and 2 min ramp to 72°C, and 2 min at 72°C. A final cycle of 72°C for 7 min was completed before the reaction mixtures were held at 4°C. Amplified DNA products were electrophoresed on 1% (w/v) agarose gels in 1× TBE buffer at 96 V. A 100 bp DNA marker (GibcoBRL, Life Technologies, Rockville, MD) was used to estimate fragment size. The gels were stained with ethidium bromide, viewed under ultraviolet light, and photographed.

Data Collection and Analyses

To evaluate the degree of random amplified polymorphic DNA (RAPD) fragment variation within and between the U.S. and Chinese soybean ancestral gene pools, DNA fragments were scored as either present (1) or absent (0). Simple matching coefficients (SMC), $S_{ij} = (a + d)/(a + b + c + d)$, were used to calculate the similarity coefficients between each pair of genotypes where a = number of fragments in common between lines; d = number of fragments absent in both lines, and b and c = number of fragments not in common between two lines. All scorable polymorphic and monomorphic fragments for each line were included to compute the similarity coefficients.

A SAS macro (Mumm and Dudley, 1995) was used to compute the similarity matrix on the basis of the SMC. Euclidean distances, $D_{ij} = (1 - S_{ij})^{1/2}$, were calculated on the basis of the similarity coefficients, and used as input into the hierarchical cluster analysis methods of UPGMA (unweighted pair group method using arithmetic average) (SAS Institute, 1989a) and Ward's minimum variance method (Ward, 1963). The TREE procedure was used to generate dendrograms. A nonhierarchical cluster analysis, VARCLUS (SAS Institute, 1989b), was also used with the original data as input to calculate the covariance matrix. UPGMA and Ward's methods are common procedures used in clustering analysis. In UPGMA, the distance between two clusters is the average distance between pairs of observations and in Ward's method the distance between two clusters is the analysis of variance sum of squares between the two clusters summed over cluster members. VARCLUS performs the disjoining clustering of variables based on a covariance matrix. There is no way of determining which procedure most accurately represents the genetic reality so multiple procedures were used to analyze the data.

The average genetic distances and standard deviations between and within the two gene pools were obtained by the MEANS procedure in SAS (SAS Institute, 1989a). Pairwise F_{st} statistics for all pairs of populations for the two gene pools were calculated using the analysis of molecular variance (AMOVA) program (Schneider et al., 1997) in which the squared Euclidean distance from cluster analysis was used as input. The significance of pairwise F_{st} values was tested by permuting the individuals between the populations by a nonparametric permutation scheme (Schneider et al., 1997).

RESULTS

RAPD Data Profile

The 35 selected RAPD primers (Thompson and Nelson, 1998) produced 261 scorable fragments of which 145 fragments were found to be polymorphic (56%). In this research, each selected primer produced an average of 7.5 fragments. Nine fragments, OPE-11700, OPF-41100, OPH-13800, OPK-10500, OPL-9750, OPN-18480, OPP-9980, OPR-71350 and OPS-14580, were found to be present in the Chinese soybean ancestral lines only. These fragments were rare among Chinese lines and each occurred in 10 or fewer ancestral lines.

Genetic distances among the 2450 pairwise combinations of U.S. and Chinese soybean ancestors ranged from 0.26 to 0.52 with a standard deviation of 0.033 and a mean genetic distance of 0.43 (Table 4). 'Arksoy' and 'Ralsoy', a selection from Arksoy, had the smallest genetic distance. 'Qin yang shui bai dou', a Chinese ancestral line from Henan province, had the overall largest genetic distance from Arksoy. Within the U.S.

gene pool, 'Lincoln', a major northern U.S. ancestor with unknown parents, and 'Haberlandt', a southern U.S. ancestor from N. Korea had the largest genetic distance (0.51). Within the Chinese gene pool, '58-161' and 'Pu dong da huang dou', from the neighboring provinces of Jiangsu and Shanghai, had the least genetic distance (0.29) and 'Hua xian da lu dou' and 'Ji ti No. 1' from Henan and Liaoning provinces, respectively, had the largest genetic distance. Between the two gene pools, 'Mukden', a major northern U.S. ancestor originally from Liaoning, had the minimum genetic distance with 'Tie jia si li huang' from Jilin province.

Although the average distances, ranges, and standard deviations between and within the two gene pools were similar, the level of diversity was slightly lower within the U.S. gene pool than within Chinese gene pool or between two gene pools (Table 4).

Genetic Patterns of the U.S. and Chinese Soybean Ancestors

Each of the clustering procedures (UPGMA, Ward's and VARCLUS) assigned the U.S. and Chinese ancestral lines to ten clusters (Table 5). There were six clusters (3, 4, 5, 6, 9, and 10) that were consistently defined by all three procedures, three clusters (1, 7, and 8) that were consistently grouped by two of three procedures, and one cluster (2) that had a diverse pattern with the three procedures (Table 5). Two of these clusters (1 and 4) contained only U.S. ancestors. Cluster 1 was predominately northern U.S. soybean ancestors from the NE region of China, whereas Cluster 4 was predominately southern U.S. soybean ancestors. Lincoln and 'Dunfield', two important northern U.S. ancestors, were assigned to Cluster 1 by two of the procedures, UPGMA and VARCLUS, but were clustered with two Chinese ancestors, 'Huang bao zhu' from Jilin and Ji ti No. 1 from Liaoning, by Ward's method. Dunfield was originally from Jilin province in NE region of China; the parents of Lincoln are unknown. Although 'S-100' in Cluster 1 is a major southern U.S. soybean ancestor, it was derived from 'Illini', which is also in Cluster 1. Illini may be one of the parents of S-100 (Thompson et al., 1998). In Cluster 4, all lines but 'Perry' were originally from Jiangsu or had one parent from Jiangsu. Perry, 'Jackson', and 'Ogden' all had parents from Japan.

Three of the clusters (5, 6, and 8) contained only Chinese soybean ancestral lines. Cluster 5 contained four ancestors from Shandong and Shanxi including 'Ju xuan 23' and 'Xin huang dou' from Shandong that were among the most important ancestors for the HHH region. Cluster 6 contained three ancestors from the three adjacent provinces of Shandong, Henan, and Jiangsu in the HHH region. Two of the three lines were among the most highly used ancestors in the HHH region. Clusters 5 and 6 included seven of the eleven major ancestors for the HHH region. Cluster 8 had three ancestors from the southern region and one ancestor from Henan, which is in the southern part of the HHH area. 'Feng xian sui dao huang' and '493-1' in Cluster 8 were grouped together by all three procedures and were the two most important ancestors for the southern region.

Table 4. Genetic distances within and between U.S. and Chinese soybean ancestral lines.

	N†	Mean distance	Standard deviation	Minimum distance	Maximum distance
U.S. ancestral lines	153	0.417	0.04	0.26	0.51
Chinese ancestral lines	496	0.429	0.03	0.29	0.50
U.S.-Chinese ancestral lines	576	0.434	0.03	0.31	0.52

† Pairwise combinations of genotypes.

The other two members of Cluster 8, 'Hua xian da lu dou' and 'Tai xing hei dou', were identified as outliers by UPGMA but assigned to this cluster by VARCLUS and Ward's methods.

The other five clusters each contained Chinese and U.S. soybean ancestral lines. Cluster 3 was predominately Chinese ancestral lines from Heilongjiang, Liaoning, and Jilin in the NE region but also contained Mukden, a major ancestor in the northern U.S. that originally came from Liaoning. Cluster 9 has two relatively minor Chinese ancestral lines from the south and HHH regions and Arksoy and Ralsoy that were minor southern U.S. ancestral lines. Cluster 10 was also a mixture of two Chinese and three U.S. ancestral lines that originated from the NE region except for Haberlandt. Although the records show that Haberlandt originally came from North Korea, it is in MG VI and is mostly in the pedigrees of cultivars developed in the southern U.S. Cluster 7 included major ancestors for three of the five major soybean-producing regions of the USA and China. 'Ai jiao zao' and 'Shanghai liu yue bei' were two of the four top contributing ancestors of southern China. 58-161 is one of the three most important ancestors of the HHH region and CNS, originally from Jiangsu, is a major ancestral line of the southern USA. This cluster also included two other ancestors from the southern region of China. Both UPGMA and VARCLUS defined Cluster 7 identically, but with Ward's 58-161, Pu dong da huang dou, and CNS were assigned to Cluster 8.

Cluster 2 was the largest group defined by UPGMA, but it was divided into three clusters by each of the Ward's and VARCLUS procedures. However, since there was very little consistency among the groupings of Ward's and VARCLUS in Cluster 2, all of these lines were put into one cluster recognizing that it is a diverse group. This cluster included six Chinese ancestral lines from the NE region of China and one U.S. ancestor 'Korean', probably from North Korea. 'Jin yuan', Huang bao zhu, and 'Zi hua No. 4', included in this cluster, were the three most important ancestral lines of the NE region of China. These three lines were assigned to different clusters by both Ward's and VARCLUS.

'Qi huang No. 1' was one of the three most important ancestors in HHH region. It was classified as an outlier in UPGMA and was not consistently grouped with any other lines by the other two procedures, so we chose not to include it in any cluster.

Genetic Relationships among Regional Populations

The clusters formed by all procedures generally reflected the geographical origin of the lines. To explore

Table 5. Cluster assignments for the U.S. and Chinese ancestral soybean lines based on three clustering methods and assigned clusters based on all of data.

Entry	Origin	Region†	UPGMA‡	WARD§	VARCLUS¶	Consensus#
A.K. (Harrow)	NE China	NUS	1	10	4	1
Illini	NE China	NUS	1	10	4	
S-100	NE China + ?	SUS	1	10	4	
Capital	NE China	NUS	1	10	4	
Dunfield	Jilin	NUS	1	9	4	
Lincoln	NE China	NUS	1	9	4	
Zi hua No. 4	Heilongjiang	NE	2	7	1	2
Korean	N. Korea	NUS	2	7	7	
Ji ti 3	Jilin	NE	2	8	1	
Jin yuan	Liaoning	NE	2	8	7	
Xiao jin huang	Jilin	NE	2	8	8	
Huang bao zhu	Jilin	NE	2	9	8	
Ji ti No. 1	Liaoning	NE	2	9	8	
Ji ti No. 5	Heilongjiang	NE	3	8	1	3
Ke shan si li jia	Heilongjiang	NE	3	8	1	
Mukden	Liaoning	NUS	3	8	1	
Tie jia si li huang	Jilin	NE	3	8	1	
Ji ti No. 2	Liaoning	NE	3	8	1	
Xiong yue xiao huang dou	Liaoning	NE	3	8	1	
Jackson	Japan-China	SUS	4	7	7	4
Roanoke	Jiangsu	SUS	4	7	7	
Ogden	Japan-China	SUS	4	7	7	
Perry	Japan-China	SUS	4	7	7	
Ju xuan 23	Shandong	HHH	5	5	6	5
Shandong si jiao qi	Shandong	HHH	5	5	6	
Jin dou No. 4	Shanxi	HHH	5	5	6	
Xin huang dou	Shandong	HHH	5	5	6	
Tong shan tian e dan	Jiangsu	HHH	6	6	5	6
Tie jiao huang	Shandong	HHH	6	6	5	
Qin yang shui bai dou	Henan	HHH	6	6	5	
Ai jiao zao	Hubei	South	7	4	3	7
Shanghai liu yue bai	Shanghai	South	7	4	3	
Hou zi mao	Hubei	South	7	4	3	
58-161	Jiangsu	HHH	7	3	3	
Pu dong da huang dou	Shanghai	South	7	3	3	
CNS	Jiangsu	SUS	7	3	3	
Nan nong 493-1	Jiangsu	South	8	3	10	8
Feng xian sui dao huang	Shanghai	South	8	3	10	
Hua xian da lu dou	Henan	HHH	O††	3	10	
Tai xing hei dou	Jiangsu	South	O	3	10	
Xiang dou No. 3	Hunan	South	9	1	9	9
Pi xian ruan tiao zhi	Jiangsu	HHH	9	1	9	
Arksoy	N. Korea	SUS	9	1	9	
Ralsoy	N. Korea	SUS	9	1	9	
Feng di huang	Jilin	NE	10	2	2	10
Richland	Jilin	NUS	10	2	2	
Da bai mei	Liaoning	NE	10	2	2	
Mandarin (Ottawa)	Heilongjiang	NUS	10	2	2	
Haberlandt	N. Korea	SUS	10	2	2	
Qi huang No. 1	Shandong	HHH	O	3	6	Outlier

† NUS = Northern U.S.; SUS = Southern U.S.; NE = Northeastern region of China; HHH = Huang Huai Hai region of China; South = Southern region of China.
‡ UPGMA = Unweighted pair group method using arithmetic average, a hierarchical cluster analysis procedure.
§ WARD = Ward's minimum variance method, a hierarchical cluster analysis procedure.
¶ VARCLUS = A nonhierarchical cluster analysis procedure.
# Consensus clusters determined by considering the groupings of all clustering procedures used.
†† Outlier in UPGMA procedure.

the genetic relationships of these geographical subpopulations, F_{st} statistics for all pairs of populations among the two U.S. and three Chinese regions were computed by the AMOVA program (Schneider et al., 1997). On the basis of the relative contribution of ancestral lines to northern and southern U.S. cultivars, northern and southern U.S. ancestral groups were defined. Lincoln, 'Mandarin (Ottawa)', 'Richland', 'A.K. (Harrow)', Dunfield, Mukden, Illini, 'Capital', and Korean were included in the northern U.S. group, and the southern U.S. group consisted of CNS, S-100, Jackson, 'Roanoke', Ogden, Haberlandt, Perry, Ralsoy, and Arksoy. Because Dunfield was in MG III, it was included in the northern group although it made a slightly larger contribution to the southern cultivars than to northern cultivars. The analyses showed that although the ancestral lines of northern and southern U.S. cultivars were generally placed in different clusters, the average genetic distances between northern and southern U.S. ancestors, and between the northern U.S. and northern Chinese ancestors were not significantly different on the basis of a nonparametric permutation test (Table 6). A relatively small significant difference between the southern U.S. and northern Chinese ancestral lines (Table 6) was noted. Highly significant differences were found between ancestral lines of northern U.S. and northeast

Table 6. Pairwise genetic distance among ancestral soybean lines from different regions of China and the USA.

	U.S.-North†	U.S.-South‡	China-Northeast§	China-HHH¶
U.S.-South	0.02			
China-Northeast	0.03	0.07**		
China-HHH	0.13**	0.10**	0.15**	
China-South#	0.16**	0.09**	0.12**	0.09**

** Significant at 0.01 probability level.
† Northern U.S. ancestral group.
‡ Southern U.S. ancestral group.
§ Northeast ancestral group of China.
¶ HHH region ancestral group of China.
# Southern ancestral group of China.

Chinese cultivars and the ancestral lines of central and southern China (Table 6).

DISCUSSION

On the basis of average genetic distance, the diversity within the Chinese and U.S. ancestral lines or between two gene pools is similar (Table 4). The pairwise F_{st} values among the regional populations (Table 6) demonstrated more specifically where differences exist. These data indicate that a relatively small genetic divergence exists between the ancestors of northern and southern U.S. cultivars, and between ancestors of northern U.S. and northern Chinese cultivars, especially when compared with the relatively large genetic distances between any of these three regions and the HHH or southern regions of China. Examining the cluster analyses may help to explain these genetic distances. In these data, the distance of northern U.S. ancestors to southern U.S. ancestors may be very small because half of the clusters that contained more than one U.S. ancestral line had both northern and southern ancestors. Perry, in MG IV, was classified as a southern ancestor although it made equal contribution to northern and southern U.S. cultivars. A similar commingling occurred between northern USA and northern China. There were three clusters that contained ancestors from the NE region of China and all contain at least one northern U.S. ancestor. In contrast, no northern U.S. or northern Chinese ancestors were found in any of the five clusters containing Chinese ancestors from the HHH or southern regions of China. Although these data (pairwise F_{st} values) may minimize the genetic distance between the northern and southern U.S. ancestors, they also highlight the genetic difference that exists between U.S. or northern Chinese cultivars and the cultivars from the HHH or southern regions of China.

The cluster analyses can help to confirm origin information, refine genetic relationships that may be assumed through origin data, and assist in defining diverse gene pools for selecting parents in cultivar improvement programs. All of the U.S. ancestral lines came from China except for three first progenies that probably had parents from both China and Japan and four ancestral lines from North Korea. One MG I line Korean, presumably from North Korea, has consistently clustered with accessions from the NE region of China in this study and in the work by Thompson et al. (1998). The other three North Korean lines, Haberlandt, Ralsoy, and Arksoy, are in MG VI, which raises some doubts about their origin. In the work of Thompson et al. (1998), these three lines clustered together with Mukden, which indicated a northern origin. In this set of lines, Haberlandt clustered with the ancestors from the NE region of China but Arksoy and Ralsoy (derived from Arksoy) clustered with ancestors from southern China, which seems inconsistent with a North Korean origin. In the work of Thompson et al. (1998), the only lines which likely would have originally come from outside NE China were U.S. ancestral lines. In this research many lines came from the HHH and southern regions of China. This change in geographical representation between these two experiments could significantly alter the clustering patterns of lines included in both studies.

Thompson and Nelson (1998) found that in U.S. soybean breeding each of the most significant crosses involving major ancestors included parents from two different genetic groups as defined by analyses of RAPD fragments. To see if that were also true with the Chinese ancestors, the major ancestors from each region were identified. They were Jin yuan, Huang bao zhu, Zi hua No. 4, 'Feng di huang', and 'Tie jia si li huang' in the NE region; 58-161, 'Tong shan tian e dan', Qi huang No. 1, Ju xuan 23, Xin huang dou, 'Tie jiao huang', and 'Pi xian ruan tiao zhi' in the HHH region; and Feng xian sui dao huang, 493-1, Ai jiao zao, and Shanghai liu yue bai in the southern region. The pedigrees of released cultivars in the three regions of China (Zhang, 1985; Wang and Wang, 1992; Hu and Tian, 1993; Cui et al., 1999) indicate that the contributions that these major ancestors made were through a few significant crosses in the NE region and through many crosses in the HHH region and south region of China.

In the NE, the contributions of the most important ancestors, Jin yuan and Huang bao zhu, were made primarily through a cross between these two lines. The contributions of other major ancestors, Feng di huang, Zi hua No. 4, and Tie jia si li huang, were made through crosses with the progenies of Jin yuan and Huang bao zhu. Jin yuan and Huang bao zhu were both assigned to Cluster 2. UPGMA put them into the same group, but the other two procedures put them in separate groups. The same occurred with Zi hua No. 4 that was also in Cluster 2. Tie jiao si li huang and Feng di huang were placed into separate clusters by all procedures.

The contributions of two of the most important ancestors, 58-161 and Tong shan tian e dan, in the HHH region were made through a cross of 58-161 and Xu dou No. 1, a selection from Tong shan tian e dan. 58-161 and Tong shan tian e dan were placed in different clusters. Zhu xu 23 and Qi huang No. 1 (also in different clusters) made their genetic contributions through a direct cross and in crosses with many other lines. The six major ancestors of the HHH regions were distributed in four clusters.

In the south, contributions of the two major ancestors, Feng xian sui dao huang and 493-1, were mostly through crosses with other cultivars and they were both put in Cluster 8. Ai jiao zao and Shanghai liu yue bei were both put in Cluster 7. Both made their genetic contribution

through a cross between these two lines, and in crosses with other lines.

The data from the Chinese ancestors were generally consistent with conclusions drawn from the U.S. ancestors. Within each of these major gene pools of soybean breeding in China, there were distinct genetic subgroups and the most significant ancestors in each gene pool were distributed in at least two different subgroups. These results provide additional evidence that the successful crosses in cultivar development often have genetically divergent parents as measured by diversity in DNA markers. Expanding genetic diversity in a breeding program could improve the performance of the resulting soybean cultivars.

The cluster analyses consistently grouped together lines with similar geographical origin and the pairwise F_{st} values obtained from the AMOVA procedure summarized these results by showing that, in general, genetic distance increased as geographical distance increased. The genetic diversities established by the cluster analyses can also define genetic divergence that exists within geographical regions and help to identify the divergent lines within gene pools. There were four clusters containing ancestors that originated from northeast China. Cluster 1 contained only U.S. ancestral lines and accounted for over 40% of the genes in northern U.S. cultivars. Cluster 10 had three U.S. ancestral cultivars that contributed nearly 30% of the genes in northern U.S. cultivars. Korean in Cluster 2 contributed less than 1% to the northern U.S. cultivars whereas this cluster had the major ancestors for the cultivars of the NE region in China. The complete separation of northern ancestors of both China and the USA. from ancestors from the HHH and south region of China indicated the large genetic differences among the lines from those regions.

This work is in general agreement with the conclusions of Cui et al. (2000b) but our data allows an actual comparison of the ancestral lines of the USA and China. Cui et al. (2000b) assumed that U.S. and Chinese ancestral lines were genetically distinct since none shared a common name. This may not be a valid assumption since all U.S. ancestral lines were named after importation into the USA and if they had a Chinese name it was generally not preserved. Our research supports the conclusion of Cui et al (2000b) that different regions of China have genetically independent gene pools but we also demonstrate the genetic relatedness of some U.S. and Chinese ancestral lines.

CONCLUSIONS

RAPD fragments exhibited a high level of efficiency for detecting the DNA polymorphism among these soybean ancestors. Most U.S. soybean ancestral lines are genetically distinct from those used in soybean breeding in China. A low level of genetic diversity was detected between southern and northern U.S. soybean ancestral lines, but the U.S. soybean ancestral lines were genetically quite distinct from the ancestral lines from HHH or southern areas in China. DNA markers such as RAPDs, combined with appropriate statistical analyses are effective tools in identifying useful genetic relationships in the absence of pedigree information and are of great value in managing and utilizing germplasm.

REFERENCES

Anonymous. 1980. The catalogue of soybean germplasm in China. Agric. Publishing House, Beijing.

Chang, R., and J. Sun. (ed.) 1991. The catalogue of soybean germplasm in China (Supplement 1). Agric. Publishing House, Beijing, China.

Cui, Z., T.E. Carter, Jr., and J.W. Burton. 2000a. Genetic base of 651 Chinese soybean cultivars released during 1923 to 1995. Crop Sci. 40:1470–1481.

Cui, Z., T.E. Carter, Jr., and J.W. Burton. 2000b. Genetic diversity patterns in Chinese soybean cultivars based on coefficient of parentage. Crop Sci. 40:1780–1793.

Cui, Z., T.E. Carter, Jr., J. Gai, J. Qui, and R.L. Nelson. 1999. Origin, description, and pedigree of Chinese soybean cultivars released from 1923 to 1995. U.S. Department of Agriculture, Agricultural Research Service, Tech. Bull. No. 1871.

Gizlice, Z., T.E. Carter, Jr., and J.W. Burton. 1994. Genetic base for North American public soybean cultivars released between 1947 and 1988. Crop Sci. 34:1143–1147.

Hu, M., and P. Tian. 1993. Annals of soybean cultivars in China (1978–1992). Agri. Publishing House, Beijing.

Keim, P., T.C. Olson, and R.C. Shoemaker. 1988. A rapid protocol for isolating soybean DNA. Soybean Genet. Newsl. 15:150–152.

Kresovich, S., W.F. Lamboy, R. Li, J. Ren, A.K. Szewc-McFadden, and S.M. Bliek. 1994. Application of molecular methods and statistical analysis for discrimination of accessions and clones of vetiver grass. Crop Sci. 34:805–809.

Mumm, R.H., and J.W. Dudley. 1995. A PC SAS computer program to generate a dissimilarity matrix for cluster analysis. Crop Sci. 35:925–927.

SAS Institute. 1989a. SAS/STAT user's guide, Version 6, 4th ed., Vol. 1. SAS Inst., Inc., Cary, NC.

SAS Institute. 1989b. SAS/STAT user's guide, Version 6, 4th ed., Vol. 2. SAS Inst., Inc., Cary, NC.

Schneider, S., J-M. Kueffer, D. Roessli, and L. Excoffier. 1997. Arlequin ver. 1.1: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sneller, C.H. 1994. Pedigree analysis of elite soybean lines. Crop Sci. 34:1515–1522.

Thompson, J.A., and R.L. Nelson. 1998. Core set of primers to evaluate genetic diversity in soybean. Crop Sci. 38:1356–1362.

Thompson, J.A., R.L. Nelson, and L.O. Vodkin. 1998. Identification of diverse soybean germplasm using RAPD markers. Crop Sci. 38:1348–1355.

Ward, J.H. 1963. Hierarchical grouping to optimize an objective function. Am. Stat. Assoc. J. 56:236–244.

Wang, L., and J. Wang. 1992. Soybean genetics and breeding. Scientific Publishing House, Beijing, China.

Zhang, Z. 1985. Annals of soybean cultivars in China. Agri. Publishing House, Beijing, China.